

# Estadística y Bioinformática

Francisco García García

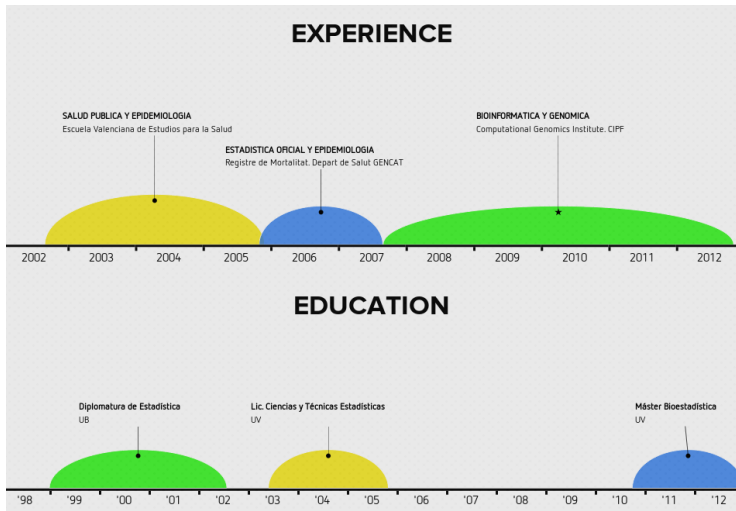
<http://bioinfo.cipf.es/fgarcia>  
*fgarcia@cipf.es*

**Universidad de Valencia**

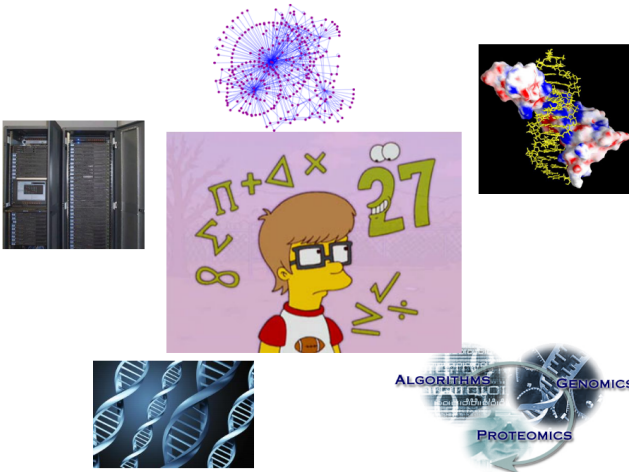
Master en Bioestadística  
Valencia, 2012

# Indice

- 1 Presentación
- 2 Bioinformática
- 3 Conclusiones



# ¿Qué hace un bioestadístico como tú en un sitio como este?

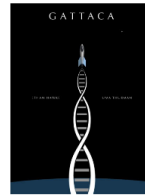


# ¿Qué es la Bioinformática?

La **Bioinformática** es el campo de la ciencia en el que la Biología, la Informática y la Estadística se unen para formar una única disciplina cuyo objetivos son permitir el descubrimiento de nuevas visiones biológicas y crear una perspectiva global que unifique los criterios que forman parte de la Biología.



# GATTACA



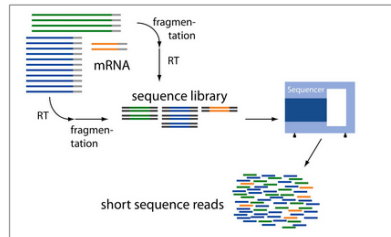
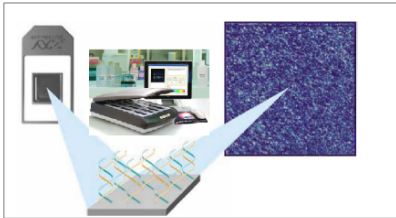
# ¿Qué actividades realiza un bioestadístico?

- 1 **Análisis de datos** genómicos
- 2 **Desarrollo de herramientas web** para análisis estadístico de datos
- 3 **Actividades formativas** sobre análisis estadístico de datos genómicos

# 1. Análisis de datos genómicos

¿Cómo obtenemos los datos?

- 1 Microarrays.
- 2 Secuenciación.

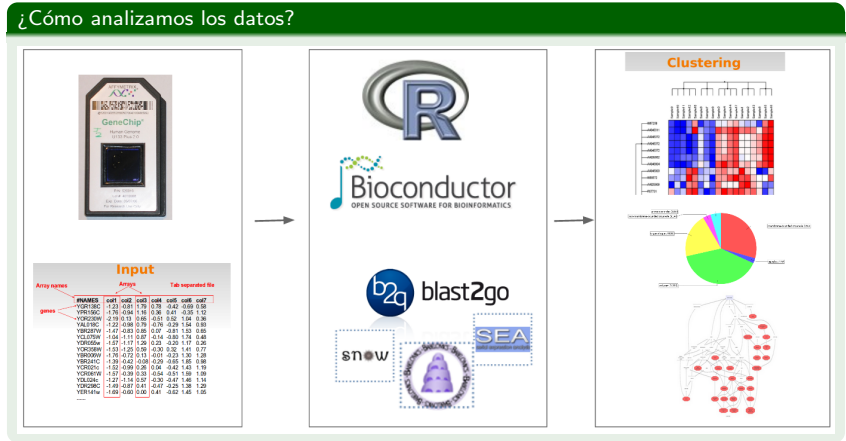


Tecnologías de alto rendimiento



# 1. Análisis de datos genómicos

## ¿Cómo analizamos los datos?



# 1. Análisis de datos genómicos

## Ejemplo: análisis de datos de expresión génica



Gene expression

- Design of experiment
- Preprocessing
- Differential expression
- Functional analysis

**Objective:** evaluating Fetal Growth Restriction (FGR)

**Design of experiment:** case-control study with paired samples.

12 Agilent's Rabbit Microarray Kit 4x44k, 1 color

# 1. Análisis de datos genómicos

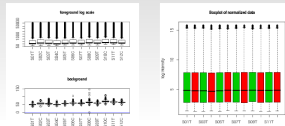
## Ejemplo: análisis de datos de expresión génica



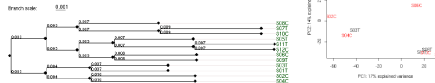
Gene expression

- Design of experiment
- Preprocessing
- Differential expression
- Functional analysis

### Raw vs. normalized data



### Exploratory analysis: clustering and PCA plot



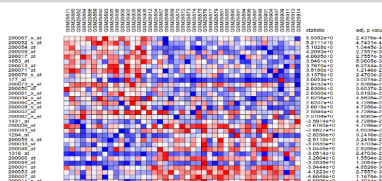
# 1. Análisis de datos genómicos

## Ejemplo: análisis de datos de expresión génica



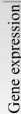
Gene expression

- Design of experiment
- Preprocessing
- Differential expression
- Functional analysis



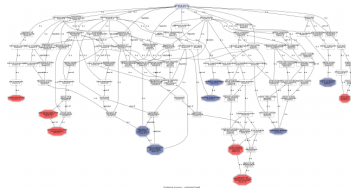
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Treatment vs control												
2	genes downregulated in treatment						genes overexpressed in control						
3		id	statistic	neg. log p-value	adj. p-value	fold-change		id	statistic	neg. log p-value	adj. p-value	fold-change	
4	1	ENSG00000100001	3.415108	0.00000007	0.040100	0.8	1	ENSG00000100001	4.56	0.00000007	0.040100	1.5	0.06
5	2	ENSG00000100002	4.858903	0.00000007	0.040100	1.08	2	ENSG00000100002	5.15	0.00000007	0.040100	1.08	0.06
6	3	ENSG00000100003	4.749778	0.00000008	0.040100	1.06	3	ENSG00000100003	4.89	0.00000008	0.040100	1.06	0.06
7	4	ENSG00000100004	4.765091	0.00000007	0.040100	1.06	4	ENSG00000100004	4.74	0.00000007	0.040100	1.06	0.06
8	5	ENSG00000100005	4.765091	0.00000007	0.040100	1.06	5	ENSG00000100005	4.74	0.00000007	0.040100	1.06	0.06
9	6	ENSG00000100006	4.765091	0.00000007	0.040100	1.06	6	ENSG00000100006	4.74	0.00000007	0.040100	1.06	0.06
10	7	ENSG00000100007	4.765091	0.00000007	0.040100	1.06	7	ENSG00000100007	4.74	0.00000007	0.040100	1.06	0.06
11	8	ENSG00000100008	4.765091	0.00000007	0.040100	1.06	8	ENSG00000100008	4.74	0.00000007	0.040100	1.06	0.06
12	9	ENSG00000100009	4.765091	0.00000007	0.040100	1.06	9	ENSG00000100009	4.74	0.00000007	0.040100	1.06	0.06
13	10	ENSG00000100010	4.765091	0.00000007	0.040100	1.06	10	ENSG00000100010	4.74	0.00000007	0.040100	1.06	0.06
14	11	ENSG00000100011	4.765091	0.00000007	0.040100	1.06	11	ENSG00000100011	4.74	0.00000007	0.040100	1.06	0.06
15	12	ENSG00000100012	4.765091	0.00000007	0.040100	1.06	12	ENSG00000100012	4.74	0.00000007	0.040100	1.06	0.06
16	13	ENSG00000100013	4.765091	0.00000007	0.040100	1.06	13	ENSG00000100013	4.74	0.00000007	0.040100	1.06	0.06
17	14	ENSG00000100014	4.765091	0.00000007	0.040100	1.06	14	ENSG00000100014	4.74	0.00000007	0.040100	1.06	0.06
18	15	ENSG00000100015	4.765091	0.00000007	0.040100	1.06	15	ENSG00000100015	4.74	0.00000007	0.040100	1.06	0.06
19	16	ENSG00000100016	4.765091	0.00000007	0.040100	1.06	16	ENSG00000100016	4.74	0.00000007	0.040100	1.06	0.06
20	17	ENSG00000100017	4.765091	0.00000007	0.040100	1.06	17	ENSG00000100017	4.74	0.00000007	0.040100	1.06	0.06
21	18	ENSG00000100018	4.765091	0.00000007	0.040100	1.06	18	ENSG00000100018	4.74	0.00000007	0.040100	1.06	0.06
22	19	ENSG00000100019	4.765091	0.00000007	0.040100	1.06	19	ENSG00000100019	4.74	0.00000007	0.040100	1.06	0.06
23	20	ENSG00000100020	4.765091	0.00000007	0.040100	1.06	20	ENSG00000100020	4.74	0.00000007	0.040100	1.06	0.06
24	21	ENSG00000100021	4.765091	0.00000007	0.040100	1.06	21	ENSG00000100021	4.74	0.00000007	0.040100	1.06	0.06
25	22	ENSG00000100022	4.765091	0.00000007	0.040100	1.06	22	ENSG00000100022	4.74	0.00000007	0.040100	1.06	0.06

## Ejemplo: análisis de datos de expresión génica



- Design of experiment
- Preprocessing
- Differential expression
- **Functional analysis**

### Summary of significant terms

[illegible]

## 2. Desarrollo de herramientas web para análisis estadístico de datos

### Babelomics



<http://babelomics.org>

## 2. Desarrollo de herramientas web para análisis estadístico de datos

### Babelomics: workflow

#### Análisis de datos de expresión



#### Regulatory elements

miRNA, CisRed  
Transcription Factor  
Binding Sites

#### Gene Ontology

Biological Process  
Molecular Function  
Cellular Component

#### KEGG pathways

**Bases de datos  
biológicas**

### 3. Actividades formativas sobre análisis de datos genómicos

50 workshops y cursos (2004-2012)



<http://bioinfo.cipf.es/courses>



<http://bioinfo.cipf.es/>

# Institute of Computational Genomics

The Department	Tools	Databases	Publications	Meetings and Courses	Services	Resources	login		
----------------	-------	-----------	--------------	----------------------	----------	-----------	-------	--	--

## Upcoming events

- Advances in the Research of Intellectual Disabilities and Autism  
Fri, 23/11/2012 (All day)
- Secuenciación Genómica en la Práctica Clínica  
Thu, 24/01/2013 (All day)
- MDA course on Next Generation Sequencing Data Analysis: from reads to candidate genes  
Mon, 11/03/2013 (All day)

[show all events](#)

## Latest news

- GenomeWeb quotes our VARIANT software  
published on 30/10/2012 - 12:23
- The Bulletin Electronique quotes our VARIANT software  
published on 30/10/2012 - 09:59
- The Valencia Board of Health, the Principe Felipe Research Center and Bull sign an agreement to establish a Chair in Computational Genomics  
published on 01/08/2012 - 16:52

[show all news](#)

## Tools usage

## Welcome

Biomedicine can only be understood in the context of genomics and with the concurrence of bioinformatics. Our department aims to tackle biomedical problems from a system's biology perspective. Following this, the general objective we seek through the main lines of research is to relate mutations to their effect at both cellular and phenotypic level trying to understand the underlying mechanism of action.

### Systems Genomics

Genes operate within an intricate network of interactions that we have only recently started to envisage. Many higher-order levels of interaction are continuously being discovered. In this scenario we are interested in developing methods and tools which can help to understand large-scale experiments from a systems biology perspective.

### Genomics of Gene Expression

Our research focusses on the dynamics and functional aspects of gene expression on the genome scale. We develop statistical approaches for the analysis of multifactorial and time-course gene expression data, focussing on their integration with functional and phenotypic data. Currently, we use Next Generation Sequencing methodologies to analyze the relationship between gene expression and genomics features. We strive to generate user-friendly tools that can routinely be used in experimental labs.

[Bioinformatics Lab](#)

Search this site:

Search

### Try our tools

## Puntos de interés

- 1 Especialización
- 2 Transversalidad
- 3 Perfil mixto

## Sugerencias

- 1 English
- 2 Equipo
- 3 Emoción

# Agradecimientos



UNIVERSITAT  
DE VALÈNCIA



PRINCIPE FELIPE  
CENTRO DE INVESTIGACION  
Computational Genomics  
Institute



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA